



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

EFFECT OF GROUPING IN GRADUATION BY OSCULATORY INTERPOLATION.

BY PERCY C. H. PAPPS.

The problem of ascertaining the rate of mortality amongst the general population is quite different from that of ascertaining the rate amongst a selected body of lives. In ascertaining the rate of mortality amongst the members of an insurance company, fraternal society, etc., it is possible to ascertain the number exposed to the risk of death at each age and the resulting deaths. Where it is desired to ascertain the rates of mortality at different ages amongst the population in the registration states, for example, it is necessary to compare as closely as possible the numbers living at each age with the resulting deaths, as in the case of a selected body of lives; but this can only be done by taking the population as shown by the census returns, made once in ten years, and comparing the numbers living at each age with the deaths at each age, as ascertained from the records of deaths, which are preferably taken from the records for a few years before and after the census.

To overcome the irregularities resulting from the data not being sufficiently extensive to give average results and to overcome errors arising in the collection and compilation of the data, graduation is necessary. In computing the rate of mortality amongst a select body of lives, an ungraduated life table may first be compiled from the ungraduated rates of mortality, and a graduated table obtained therefrom. In handling population statistics it is usual to graduate both the population and the deaths, and then to derive what is a graduated rate by dividing the graduated deaths by the graduated population statistics. It must be remembered that in handling population statistics much more extensive errors in the collection of the data have to be handled than where the rates of mortality are to be determined from the record of an insurance company or a fraternal society.

In the June 1910 QUARTERLY PUBLICATIONS of the Ameri-

can Statistical Association, pages 86 to 109, will be found an account of the graduation of the data derived from the twelfth census of the registration states by Professor Glover. The graduation is made by Osculatory Interpolation and the table on page 100 shows that the well recognized errors due to the tendency to overstate the population at the quinquennial ages ending in 0 and 5 are distributed in each case over the succeeding four ages. For example, the sums of the graduated and ungraduated population for ages 35 to 39 inclusive are identical so that the overstatement of the population at age 35 is spread over ages 36 to 39.

It is proposed to write the graduated values in terms of the ungraduated values for the formula actually used as well as for four similar formulae derived from grouping ages ending in 1 to 5, 2 to 6, 3 to 7 and 4 to 8.

The graduation is made by first summing the numbers in the column showing the population from the bottom up, and then operating on this summation column which gives the population at each age and all higher ages. The deaths were graduated in a similar manner.

Now, let u_x, u_{x+1}, u_{x+2} , etc., be terms of the original series,

$$U_x = \sum_x^{\omega} u_x$$

$$\Delta U_x = U_{x+5} - U_x$$

and

$$\delta U_x = U_{x+1} - U_x = -u_x$$

Then, writing the leading quinquennial differences of U_x in terms of the original series, the results shown in the following table are arrived at, where the coefficients are shown in the columns and the expressions to which they apply in the headings of the columns.

TABLE A.

	$\Sigma_{x+20}^{x+24} u_x$	$\Sigma_{x+15}^{x+19} u_x$	$\Sigma_{x+10}^{x+14} u_x$	$\Sigma_{x+5}^{x+9} u_x$	$\Sigma_x^{x+4} u_x$
ΔU_x					-1
$\Delta^2 U_x$				-1	+1
$\Delta^3 U_x$			-1	+2	-1
$\Delta^4 U_x$		-1	+3	-3	+1
$\Delta^5 U_x$	-1	+4	-6	+4	-1

On page 93 Professor Glover gives a table showing the coefficients of the expressions for the subdivided differences in terms of the differences for intervals of five ages. The table in terms of the present notation is as follows:

TABLE B.

	ΔU_{x-10}	$\Delta^2 U_{x-10}$	$\Delta^3 U_{x-10}$	$\Delta^4 U_{x-10}$	$\Delta^5 U_{x-10}$
δU_x	.2	.32	.088	-.0176	.0016
$\delta^2 U_x$.04	.048	.0016	.0048
$\delta^3 U_x$.008	.0064	-.0048
$\delta^4 U_x$.0016	-.0032
$\delta^5 U_x$.0080

By means of Tables A and B the leading yearly differences of U_x may be computed in terms of the original series. The results are shown in Table C.

TABLE C.

	$\Sigma_{x+10}^{x+14} u_x$	$\Sigma_{x+5}^{x+9} u_x$	$\Sigma_x^{x+4} u_x$	$\Sigma_{x-5}^{x-1} u_x$	$\Sigma_{x-10}^{x-6} u_x$
δU_x	-.0016	.0240	-.1504	-.0843	.0128
$\delta^2 U_x$	-.0048	.0176	-.0720	.0704	-.0112
$\delta^3 U_x$	+.0048	-.0256	+.0400	-.0224	+.0032
$\delta^4 U_x$	+.0032	-.0144	+.0240	-.0176	+.0048
$\delta^5 U_x$	-.0080	+.0320	-.0480	+.0320	-.0080

The first line in Table C gives one formula for ascertaining a graduated value of u_x , for $u_x = -\delta U_x$ and by writing δU_{x+1} , δU_{x+2} , etc., in terms of the differences shown in Table C, four other formulae may be found. By the usual formula the values in Table D are found.

TABLE D.

	δU_x	$\delta^2 U_x$	$\delta^3 U_x$	$\delta^4 U_x$	$\delta^5 U_x$
δU_x	+1				
δU_{x+1}	+1	+1			
δU_{x+2}	+1	+2	+1		
δU_{x+3}	+1	+3	+3	+1	
δU_{x+4}	+1	+4	+6	+4	+1

By means of Tables C and D the following formulae are obtained, the one first derived in Table C being repeated.

TABLE E.

	$\Sigma_{x+10}^{x+14} u_x$	$\Sigma_{x+5}^{x+9} u_x$	$\Sigma_x^{x+4} u_x$	$\Sigma_{x-5}^{x-1} u_x$	$\Sigma_{x-10}^{x-6} u_x$
δU_x	-.0016	+.0240	-.1504	-.0848	+.0128
δU_{x+1}	-.0064	+.0416	-.2224	-.0144	+.0016
δU_{x+2}	-.0064	+.0336	-.2544	+.0336	-.0064
δU_{x+3}	+.0016	-.0144	-.2224	+.0416	-.0064
δU_{x+4}	+.0128	-.0848	-.1504	+.0240	-.0016

It will be noticed that the first and last and second and fourth formulae in Table E have similar coefficients but in reverse order. The coefficients in the third formula are symmetrical.

Now, any value such as δU_y may be derived as shown in the following table:

TABLE F.

δU_y Derived from	Age Groupings						
δU_y	y	to	$y+4$,	$y+5$	to	$y+9$	etc.
$\delta U_{y-1}+1$	$y-1$	"	$y+3$,	$y+4$	"	$y+8$	"
$\delta U_{y-2}+2$	$y-2$	"	$y+2$,	$y+3$	"	$y+7$	"
$\delta U_{y-3}+3$	$y-3$	"	$y+1$,	$y+2$	"	$y+6$	"
$\delta U_{y-4}+4$	$y-4$	"	y ,	$y+1$	"	$y+5$	"

It is interesting to notice that if one fifth of the sum of the five formulae shown in Table E be taken, we obtain the twenty-nine term Osculatory Interpolation formula given by Mr. George King on page 559 of Volume 41 of the Journal of the Institute of Actuaries. The method of arriving at this formula is shown in the following table, and the table gives a ready comparison of the different formulae in terms of the original series. In this table the different formulae run vertically instead of horizontally as in the previous tables.

TABLE G.

	δU_y	$\delta U_{y-1}+1$	$\delta U_{y-2}+2$	$\delta U_{y-3}+3$	$\delta U_{y-4}+4$	Total	1/5 Total
$ux-14$					-.0016	-.0016	-.00032
$ux-13$				-.0064	-.0016	-.0080	-.00160
$ux-12$			-.0064	-.0064	-.0016	-.0144	-.00288
$ux-11$		+.0016	-.0064	-.0064	-.0016	-.0128	-.00256
$ux-10$	+.0128	+.0016	-.0064	-.0064	-.0016	0	0
$ux-9$	+.0128	+.0016	-.0064	-.0064	+.0240	+.0256	+.00512
$ux-8$	+.0128	+.0016	-.0064	+.0416	+.0240	+.0736	+.01472
$ux-7$	+.0128	+.0016	+.0336	+.0416	+.0240	+.1136	+.02272
$ux-6$	+.0128	-.0144	+.0336	+.0416	+.0240	+.0976	+.01952
$ux-5$	-.0848	-.0144	+.0336	+.0416	+.0240	0	0
$ux-4$	-.0848	-.0144	+.0336	+.0416	-.1504	-.1744	-.03488
$ux-3$	-.0848	-.0144	+.0336	-.2544	-.1504	-.4384	-.08768
$ux-2$	-.0848	-.0144	-.2544	-.2224	-.1504	-.7264	-.14528
$ux-1$	-.0848	-.2224	-.2544	-.2224	-.1504	-.9344	-.18688
ux	-.1504	-.2224	-.2544	-.2224	-.1504	-1.0000	-.20000
$ux+1$	-.1504	-.2224	-.2544	-.2224	-.0848	-.9344	-.18688
$ux+2$	-.1504	-.2224	-.2544	-.0144	-.0848	-.7264	-.14528
$ux+3$	-.1504	-.2224	+.0336	-.0144	-.0848	-.4384	-.08768
$ux+4$	-.1504	+.0416	+.0336	-.0144	-.0848	-.1744	-.03488
$ux+5$	+.0240	+.0416	+.0336	-.0144	-.0848	0	0
$ux+6$	+.0240	+.0416	+.0336	-.0144	+.0128	+.0976	+.01952
$ux+7$	+.0240	+.0416	+.0336	+.0016	+.0128	+.1136	+.02272
$ux+8$	+.0240	+.0416	-.0064	+.0016	+.0128	+.0736	+.01472
$ux+9$	+.0240	-.0064	-.0064	+.0016	+.0128	+.0256	+.00512
$ux+10$	-.0016	-.0064	-.0064	+.0016	+.0128	0	0
$ux+11$	-.0016	-.0064	-.0064	+.0016		-.0128	-.00256
$ux+12$	-.0016	-.0064	-.0064			-.0144	-.00288
$ux+13$	-.0016	-.0064				-.0080	-.00160
$ux+14$	-.0016					-.0016	-.00032

It will be noticed that the formulae in Table G give the values of δU_y or $-u_y$, so that to obtain the values of u_y or u_x the signs must be changed.

The first five formulae in Table G may be more readily applied if the column of U_x be first computed and the five formulae written in terms of U_x instead of u_x . In Table H the formulae are written in terms of U_x and the signs are changed so that u_x instead of $-u_x$ is obtained directly from the formulae.

TABLE H.

	δU_y	$\delta U_{y-1}+1$	$\delta U_{y-2}+2$	$\delta U_{y-3}+3$	$\delta U_{y-4}+4$
U_{x-14}					+ .0016
U_{x-13}				+ .0064	
U_{x-12}			+ .0064		
U_{x-11}					
U_{x-10}	-.0128	-.0016			
U_{x-9}					-.0256
U_{x-8}				-.0480	
U_{x-7}			-.0400		
U_{x-6}		.0160			
U_{x-5}	+.0976				
U_{x-4}					+ .1744
U_{x-3}				+.2640	
U_{x-2}					
U_{x-1}			.2880		
U_x	+.0656	.2080			
U_{x+1}					-.0656
U_{x+2}				-.2080	
U_{x+3}					
U_{x+4}			-.2880		
U_{x+5}	-.1744				
U_{x+6}					-.0976
U_{x+7}				-.0160	
U_{x+8}					
U_{x+9}			.0400		
U_{x+10}	+.0256				
U_{x+11}					+.0128
U_{x+12}					
U_{x+13}					
U_{x+14}					
U_{x+15}	-.0016				

For the population the T_x column corresponds with the U_x of the above table. By setting the values of T_x on the multiplying machine and multiplying by the coefficients shown in Table H, without clearing the product holes, the graduated value of u_x , in this case L_x , are obtained directly; care being taken to see that the machine is set for multiplication when the coefficient is positive and for division when the coefficient is negative. If one of these short formulae is to be used, the graduated values can be more directly and rapidly obtained in this manner than by building up the differences and subdivided differences shown by Professor Glover.

In the following tables are shown graduated values of L_x , d_x and q_x according to the five formulae. The average of

these values is shown which corresponds to the values that might be obtained directly by Mr. King's twenty-nine term Osculatory Interpolation formula, which may be applied by the summation method. The values are shown for ages 30 to 40 inclusive. Tables I, J and K give the values of L_x , d_x and q_x ; the latter derived from the formula used by Professor Glover, namely

$$q_x = \frac{d_x}{L_x + \frac{1}{2}d_x}$$

TABLE I—VALUES OF L_x .

x	x to $x+4$ $x+5$ " $x+9$	$x-1$ to $x+3$ $x+4$ " $x+8$	$x-2$ to $x+2$ $x+3$ " $x+7$	$x-3$ to $x+1$ $x+2$ " $x+6$	$x-4$ to x $x+1$ " $x+5$	Average Values
30	173,618	172,937	177,940	176,353	177,033	175,576
31	169,201	169,372	168,500	173,088	171,805	170,393
32	166,978	160,405	165,600	165,228	167,208	165,084
33	161,093	161,563	154,471	162,809	162,653	160,518
34	159,571	154,160	157,264	153,401	160,559	156,991
35	157,926	156,121	149,430	154,912	155,210	154,720
36	156,259	155,128	152,851	148,350	153,658	153,249
37	152,032	157,691	151,846	149,880	149,366	152,163
38	149,889	150,573	156,234	147,839	146,987	150,304
39	144,042	150,934	147,535	150,025	143,313	147,170
40	138,832	141,327	148,912	141,978	140,883	142,386

TABLE J—VALUES OF d_x .

x	x to $x+4$ $x+5$ " $x+9$	$x-1$ to $x+3$ $x+4$ " $x+8$	$x-2$ to $x+2$ $x+3$ " $x+7$	$x-3$ to $x+1$ $x+2$ " $x+6$	$x-4$ to x $x+1$ " $x+5$	Average Values
30	1,445	1,452	1,483	1,453	1,487	1,464
31	1,468	1,486	1,448	1,479	1,473	1,461
32	1,488	1,441	1,437	1,461	1,471	1,460
33	1,459	1,499	1,434	1,455	1,484	1,466
34	1,501	1,440	1,510	1,461	1,484	1,479
35	1,509	1,517	1,440	1,520	1,508	1,499
36	1,549	1,584	1,527	1,470	1,529	1,522
37	1,538	1,595	1,548	1,532	1,516	1,546
38	1,559	1,546	1,614	1,546	1,551	1,559
39	1,531	1,609	1,546	1,591	1,533	1,562
40	1,522	1,538	1,626	1,533	1,543	1,552

TABLE K—VALUES OF q_x .

x	x to $x+4$ $x+5$ " $x+9$	$x-1$ to $x+3$ $x+4$ " $x+8$	$x-2$ to $x+2$ $x+3$ " $x+7$	$x-3$ to $x+1$ $x+2$ " $x+6$	$x-4$ to x $x+1$ " $x+5$	Average Values
30	.008288	.008361	.008300	.008205	.008364	.008304
31	.008639	<i>.008443</i>	.008557	.008508	.008537	.008538
32	.008872	.008943	<i>.008640</i>	.008803	.008759	.008805
33	.009016	.009235	.009240	<i>.008897</i>	.009082	.009091
34	.009362	.009298	.009556	.009479	<i>.009200</i>	.009377
35	<i>.009510</i>	.009670	.009590	.009764	.009669	.009642
36	.009864	<i>.009840</i>	.009940	.009860	.009901	.009882
37	.010065	.010064	<i>.010143</i>	.010170	.010098	.010109
38	.010347	.010215	.010278	<i>.010403</i>	.010362	.010319
39	.010573	.010604	.010424	.010549	<i>.010640</i>	.010558
40	<i>.010903</i>	.010824	.010860	.010739	.010893	.010841

It will be noticed that Professor Glover used all five of the separate formulae, the figures corresponding to those derived by him being printed in italics in Tables I, J and K. In a few places slight differences may be noted between the figures given above and Professor Glover's figures. These are probably due to dropping decimals in the method used by Professor Glover in obtaining his figures.

The values in Table K may be rearranged so that the values derived from grouping ages 30–34, 35–39, etc., are in one column and other groupings in the other columns. By this table a comparison of the application of the method used by Professor Glover to the different age groupings can be made more readily. It may be well to bear in mind that no matter what grouping is used, the totals of the graduated and ungraduated values will agree within the five age groups. The values of q_x are rearranged in Table L.

TABLE L—VALUES OF q_x .

x	0-4 5-9	1-5 6-0	2-6 7-1	3-7 8-2	4-8 9-3	Average Values
30	.008288	.008364	.008205	.008300	.008361	.008304
31	.008443	.008639	.008537	.008508	.008557	.008538
32	.008640	.008943	.008872	.008759	.008803	.008805
33	.008897	.009240	.009235	.009016	.009082	.009091
34	.009200	.009479	.009556	.009298	.009362	.009377
35	.009510	.009669	.009764	.009590	.009670	.009642
36	.009840	.009864	.009901	.009860	.009940	.009882
37	.010143	.010064	.010065	.010098	.010170	.010109
38	.010403	.010278	.010215	.010347	.010362	.010319
39	.010640	.010549	.010424	.010604	.010573	.010558
40	.010903	.010893	.010739	.010860	.010824	.010841

As a test of the smoothness of the graduated values, the third differences are shown in Table M, but carried to one more decimal place than in Table X of Professor Glover's article.

TABLE M—VALUES OF $10^6 \Delta^3 q_x$.

x	0-4 5-9	1-5 6-0	2-6 7-1	3-7 8-2	4-8 9-3	Average Values
30	+18	-36	+25	-37	-17	-14
31	-14	-51	-70	+19	-32	-19
32	-39	+ 9	-71	-15	+27	-21
33	+13	+54	+42	-32	-66	-4
34	-47	0	+98	-10	- 2	+12
35	-16	+ 9	-41	+43	+ 2	- 4
36	+20	+43	+73	- 3	+57	+46
37	+49	+16	+47	- 9	+21	+15
Totals*	216	218	467	168	224	135

* Irrespective of sign.

As might be expected, the sum of the third differences irrespective of sign is smallest for the powerful twenty-nine term Osculatory Interpolation formula, but of the five short formulae the smallest sum is for the grouping where the quinquennial ages ending in 0 and 5 are the center of the groups. With a sufficient amount of data to allow full play for the operation of the law of averages, we would expect very little unevenness in the values of q_x between ages 30 and 40, were

it not for the misstatement of ages. The tendency to give an age ending in 0 or 5 is very strongly shown in the data to be graduated, and it seems not unreasonable to suppose that the graduation which will most satisfactorily remove this error will be the one showing the smoothest graduated values; provided, however, that the graduation formula does not remove any characteristics of the data, not due to error, which should be retained. So far as the limited test shows, the grouping of the ages with those ending in 0 and 5 as the center of the groups would seem to be more satisfactory than the groupings used by Professor Glover. In the first case the errors are spread over the younger as well as the older ages, while Professor Glover's grouping throws the errors entirely over the older ages of each quinquennial group.

In Table N are shown one million times the differences between the values of q_x by the twenty-nine term formula and by the shorter formulae. The positive and negative differences are added and the net differences and total differences irrespective of sign are shown at the foot of the table.

TABLE N.

ONE MILLION TIMES DIFFERENCES IN VALUES OF q_x BY TWENTY-NINE TERM AND SHORTER FORMULAE.

x	0-4 5-9	1-5 6-0	2-6 7-1	3-7 8-2	4-8 9-3
30	- 16	+ 60	- 99	- 4	+57
31	- 95	+101	- 1	-30	+19
32	-165	+138	+ 67	-46	- 2
33	-194	+149	+144	-75	- 9
34	-177	+102	+179	-79	-15
35	-132	+ 27	+122	-52	+28
36	- 42	- 18	+ 19	-22	+58
37	+ 34	- 45	- 44	-11	+61
38	+ 84	- 41	-104	+28	+43
39	+ 82	- 9	-134	+46	+15
40	+ 62	+ 52	-102	+19	-17
Positive	262	629	531	93	281
Negative	821	113	484	319	43
Net	-559	+516	+ 47	-226	+238
Total*	1083	742	1015	412	324

* Irrespective of sign.

Table N shows that the grouping used by Professor Glover produces values further from those given by the twenty-nine term formula than any of the other groupings, whether measured by the net difference of the eleven values or by the sum of the differences irrespective of sign. While the grouping 2 to 6, 7 to 1 gives the closest agreement in the aggregate, the individual differences are large. The next in order is the grouping having central ages ending in 0 or 5, and in this case the differences are not large. The grouping 4 to 8, 9 to 3 gives nearly as good results according to Table N, but Table M shows that the graduated values of q_x do not run as smoothly as with groupings 8 to 2, 3 to 7, where the central ages end in 0 or 5.

In conclusion it may be said that in using Professor Glover's plan of graduation the groupings to be used should be carefully considered. It is recognized that the foregoing study takes into account only eleven ages, but they are important ages and were chosen for the reason that marked changes in the progression of the rate of mortality are not looked for between ages 30 and 40. This paper is submitted with the thought that the points herein raised, if of value, may be borne in mind when the time comes to graduate the next Census Table.